Sustainable Computing Project Proposal
Eddy Abban

3-24-2025

Tufts University

Motivation

Machine learning workloads are becoming increasingly common, especially for for edge devices, many of which now contain machine learning accelerators. These accelerators are a part of the larger SoC due to the size restrictions of many of these devices. However, there has still been development in off chip accelerators for edge devices. For example, Google has created Coral USB Accelerator which is able to accelerate machine learning workloads on edge devices such as a Raspberry Pi. The sustainability of such ASIC based accelerators had not been thoroughly compared to that of an FPGA in edge computer scenarios. While the ASICs offer higher energy efficiency their are a fixed function and cannot be reprogrammed. FPGAs on the other hand are reconfigurable allowing for longer usability and adaptability, however FPGAs generally require more power, larger die sizes, and a more intensive design process. This project aims to evaluate the sustainability trade-offs between these two hardware configurations for edge-based machine learning workloads.

Background

Various hardware accelerators exist to enhance machine learning performance beyond traditional CPUs. For example, many modern systems use GPUs in order to accelerate machine learning. Even more specialized workloads utilize ASICs such as TPUs which are solely optimized for machine learning tasks.

FPGAs present an alternative with unique trade-offs. Unlike ASICs which are fixed upon fabrication, FPGAs can be reprogrammed to adapt to different workloads. This allows for FPGAs to keep up with rapid changes and developments and extends their life span. FPGAs however due to their reconfigurable nature come with some drawbacks. They require more die space and also require more electricity to operate. Also generating the layout to be placed on an FPGA is also an extremely intensive task. All of these factors must be taken into account when assessing each framework for sustainability.

Related work

Green FPGA

This work introduces a novel tool designed to evaluate the environmental sustainability of FPGAs by comprehensively modeling their carbon footprint across their entire lifecycle, from design and manufacturing to operation, disposal, and recycling. Recognizing the growing ecological impact of computing, the study focuses on FPGAs due to their reconfigurability, which offers potential sustainability advantages over ASICs. GreenFPGA analyzes scenarios where the benefits of FPGA reconfigurability,

such as extended lifespan and reuse across multiple applications, can offset their higher embodied carbon costs. Experimental results demonstrate that FPGAs exhibit lower carbon footprints than ASICs in specific conditions, including low-volume applications, short application lifespans (below 1.6 years), and when used for multiple applications (over five) or with lower application volumes (under 2 million for iso-performance). By providing a detailed model for design carbon footprint and incorporating unique aspects of FPGA-based computing, GreenFPGA offers a valuable tool for assessing and promoting more environmentally sustainable hardware acceleration solutions. This paper provides the framework and some of the tools that we will be using to analyze the sustainability of FPGAs.

ACT

The research paper "ACT: Designing Sustainable Computer Systems With An Architectural Carbon Modeling Tool" addresses the critical shift in computing's carbon footprint from operational emissions to embodied emissions, primarily due to hardware manufacturing. Recognizing the lack of architectural modeling tools to quantify and optimize this end-to-end carbon footprint, the authors introduce ACT, a novel framework designed to facilitate sustainability-driven design space exploration. ACT models both operational and embodied emissions, offering a detailed breakdown of the latter for components such as SoCs, DRAM, and storage, using data from semiconductor fabs and industry environmental reports. The paper highlights that optimizing for carbon yields distinct design choices compared to traditional performance and efficiency optimization, and introduces use-case-dependent carbon optimization metrics to guide this process. Employing the "Reduce, Reuse, Recycle" framework, the authors demonstrate the tool's capabilities by analyzing the trade-offs between general-purpose and specialized hardware, optimizing accelerators for minimal carbon footprint while meeting performance targets, and exploring the benefits of extending hardware lifespan through improved reliability. Furthermore, the paper critiques existing life cycle analysis (LCA) tools for their coarse-grained data and limited IC footprint information, advocating for a more granular approach to hardware design. Ultimately, the authors issue a call to action for computer architects to prioritize sustainability as a primary design consideration, and make the ACT model and configurable parameters available to the community to foster further research in this critical area. This paper provides the framework and tools for analyzing the sustainability of the ASICs.

Planned work

The goal of this project is to compare the sustainability of ASIC and FPGA based running machine learning by analyzing both their embodied and operational carbon footprints.

1. **Hardware:**

    a. FPGA: Digilent Basys 3 Artix-7 FPGA Trainer Board
    b. ASIC Accelerator: Coral USB Accelerator
2. **Machine Learning Task:**
    a. Image classification using the **Fashion-MNIST** dataset
3. **Methodology**
    a. Apply the **ACT** framework to estimate the embodied carbon footprint of each device.
    b. Utilize data from manufacturing reports and previous LCA studies (such as GreenFPGA).
4. **Operation carbon Footprint Analysis**
    a. Run the classification model on both devices and measure real-time power consumption
    b. Use local electricity generation data (Boston's carbon emissions per watt) to estimate the operational carbon footprint
5. **Model Optimization & Comparison**
    a. Implement model quantization using TensorFlow and use HLS4ML to generate FPGA-compatible designs.
    b. Compare
        i. FPGA running a quantized model.
        ii. Coral running a non-quantized model.
        iii. Coral running a quantized model.
    c. Evaluate the trade-offs between energy efficiency, performance, and overall sustainability.
6. **Note:**
    a. I will be focus creating the initial FPGA designs of the machine learning workloads since that is most likely the hardest aspect of this project. The coral aspect of this project may not happen if there is not sufficient time.

Citations

1. HLS4ML
2. ACT
3. fashion mnist
4. Green FPGA
5. Coral USB Accelerator
6. Digilent Basys 3 Artix-7 FPGA Trainer Board